

## 社会における知の深い分析と有効活用

## ■概要

データ駆動知能システム研究センターでは、研究内容は、大規模なテキストデータを対象とする自然言語処理を中心とし、ネット等にテキストとして流布している、社会における知、すなわち社会知を意味的に深く分析し、有効活用できる技術を開発することである。より具体的には、前身である情報分析研究室が開発、一般公開した大規模情報分析システムWISDOM X、耐災害ICT研究センターと共同で開発した対災害情報分析システムDISAANA等で使われている技術を発展させ、よりシステムが自律的に分析を行える枠組みを開発し、また、その分析の結果を、一般市民を含め、多くのユーザにわかりやすく提供できる技術の開発を目指す。平成29年度は、社会知の有効活用を目指し、Web上における大量の知識を活用して多様なトピックに関する対話を行う次世代音声対話システムプロトタイプ「WEKDA（ウェクダ）」の研究開発を平成28年度（平成28年度では「WISDOMくん」と呼称していた）に引き続いて実施し、さらにはWEKDAでの活用を意識して、WISDOM Xでは長文で回答していた「なぜ」型質問の回答を対話中で自然な短い文に深層学習を用いて要約する技術や、テキスト中に書かれている問題の解決法をやはり深層学習を用いて要約する技術等を開発した。また、WISDOM Xの質問応答技術の精度向上や新規アルゴリズムの開発を深層学習を用いて行ったほか、DISAANAや災害状況要約システムD-SUMMの拡張や防災訓練等への参加を行った。特に大分県では、防災訓練実施後の7月の九州北部豪雨の際に、県庁が実際にDISAANA、D-SUMMを活用し、鉄橋の流出を他機関に先駆けて検知するなど県の災害対応において役立った。さらに、深層学習の多用が今後研究で必要なことから、WISDOM X等を開発するために独自開発し、ミドルウェアRaSCに深層学習のフレームワークを統合し、多数の深層学習モジュールからなるWEKDAの応答速度を向上させた。さらにはGPGPUクラスタに独自のバジスケジューラを導入し、GPGPUの利用効率の改善を図った。これらDISAANA、D-SUMMについては3.10.5耐災害ICT研究センターの項で詳述し、本稿では対話システムWEKDAを中心に説明する。

## ■平成29年度の成果

次世代音声対話システムプロトタイプ「WEKDA」は、図1に示すように人と音声で対話をする対話システムである。既に複数の民間企業から発売され、実用化が進められている対話システムは、その多くがあらかじめ人手によって記述した何らかのルールに基づいて対話を行うシステムであり、特定のタイプの命令を入力されるとその命令を実行するといった動作は容易であるが、システム開発時に想定していないタイプの命令が入力された場合には、対応が難しい。一方で、近年注目を集めている深層学習をベースとする対話システムは、大量の対話データから先述の対話のルールに相当するものを学習し、入力に対して適切と思われる出力をユーザに返すものである。しかしながら、学習データを増やしていくと、相槌等の内容の薄い出力しか返さず、そのままでは自然な対話の実現は困難であることが知られている。

WEKDAの対話例を図2、3に示す。これらが示すように、現状WEKDAはiPS細胞のような高度に科学的な対話から、観光のようなやわらかいトピックの対話に至るまでの広範な対話において、一定の割合で適切な応答を行っている。WEKDAはその背後に、Web40億ページの情報から質問への回答を提示する大規模Web情報分析システムWISDOM Xが控えており、入力文（例：「iPS細胞ってすごいね？」）が与えられると、その入力文からユーザに提示すべき情報を問う質問（例：「iPS細胞で何を見る？」）を深層学習で生成する。この質問はWISDOM X



図1 次世代音声対話システムWEKDAとの対話の様子



図2 WEKDAの対話例1



図3 WEKDAの対話例2

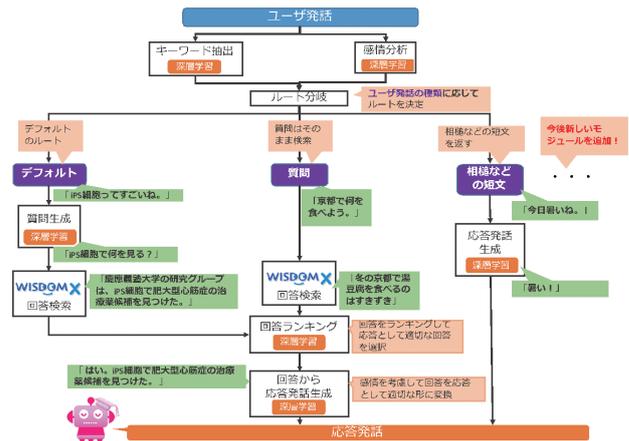


図4 対話システムWEKDAのアーキテクチャ

に与えられ、その回答検索の結果からシステムの応答（例：「iPS細胞で肥大型心筋症の治療薬候補を見つけた。」）を生成する。この仕組みの大枠は前年度に開発されたものであるが、平成29年度は、回答検索の結果からより適切と思われる応答を生成するために、深層学習を用いて回答検索の結果をランキングする機構並びに回答から適切な応答を機械翻訳に類似する深層学習手法で生成する機能を実現し、より適切な応答が得られるようになった。図4にWEKDAのアーキテクチャを示すが、これらの機構を含めて多数の深層学習モジュールの組み合わせで実現された対話システムとなっており、また、質問が入力された場合は直接WISDOM Xに問い合わせをして回答を提示する仕組みとなっている。

現在、WEKDAで対応できる質問は、「何で地球温暖化を解決する？」のように名詞一語で回答する「何」型質問、「地球温暖化が進むとどうなる？」といった「どうなる」型質問に限られているが、本来、WISDOM Xでは「なぜ日本はデフレに陥ったか？」のように物事の理由を尋ねる「なぜ」型質問にも回答ができる。しかし、比較的長い文章で提示されるため応答としては不適切である。今年度こうした長い文を要約する技術を、深層学習を用いて開発し、前出の質問に対して「少子高齢化が進んでいるため」といった短い回答を返すことが可能になっている。

類似する研究として、「PCがハングした。しょうがないのでリポートした。」のように、問題とその解決策が記述されているテキストから、「PCがハングした」を問題、「リポートすること」を解決策としてそれぞれ、要約、

抽出する技術も深層学習を用いて新たに開発している。さらに、WISDOM X上で現在一般公開している「なぜ」型質問応答の精度改善も新たなニューラルネットワークアーキテクチャを開発、考案することで、1位にランキングされた回答の精度が昨年度に比べて6%改善した。（ただし、公開されているWISDOM Xへの組み込みは未実施。）また、現在WISDOM Xには実装されていない何らかの行為の方法を問う「How」型質問応答の手法も新たなネットワークアーキテクチャを用いて実現し、1位にランキングされた回答の精度が従来法に比べて27%向上した。

最後に以上の技術開発を支える計算基盤として、当センターで開発し、これまでWISDOM X等の大規模なソフトウェアを大規模クラスタ上で高並列かつ高速に稼働させるために利用して来たミドルウェアRaSCに深層学習フレームワークを統合した。これにより、深層学習の固まりとも言えるWEKDAをローコストで高速に稼働させることに成功している。また、同様に研究時のGPGPUの利用効率の向上を図るため、独自バッチスケジューラを開発し、センター内のGPGPU利用の高効率化を実現した。

今後は、質問応答・要約技術等の基盤的自然言語処理技術の高度化、計算基盤の高度化を進めるとともに、WEKDAの対話の適切さを向上させ、また、WEKDAに対話の目標を持たせる、あるいは過去の対話履歴等を参照した文脈を意識した対話を行わせる等の高度な対話技術の開発を進める予定である。